

# A change-point approach for the identification of financial extreme regimes

Chiara Lattanzi<sup>1</sup> and Manuele Leonelli<sup>2</sup>

<sup>1</sup>*Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italy, chiara.lattanzi3@studio.unibo.it*

<sup>2</sup>*School of Human Sciences and Technology, IE University, Madrid, Spain, manuele.leonelli@ie.edu*

**Abstract.** Inference over tails is usually performed by fitting an appropriate limiting distribution over observations that exceed a fixed threshold. However, the choice of such threshold is critical and can affect the inferential results. Extreme value mixture models have been defined to estimate the threshold using the full dataset and to give accurate tail estimates. Such models assume that the tail behavior is constant for all observations. However, the extreme behavior of financial returns often changes considerably in time and such changes occur by sudden shocks of the market. Here the extreme value mixture model class is extended to formally take into account distributional extreme change-points, by allowing for the presence of regime-dependent parameters modelling the tail of the distribution. This extension formally uses the full dataset to both estimate the thresholds and the extreme change-point locations, giving uncertainty measures for both quantities. Estimation of functions of interest in extreme value analyses is performed via MCMC algorithms. Our approach is evaluated through a series of simulations, applied to real data sets and assessed against competing approaches. Evidence demonstrates that the inclusion of different extreme regimes outperforms both static and dynamic competing approaches in financial applications.

## 1 Introduction

The financial market is characterized by periods of turbulence where extreme events shock the system, potentially leading to huge profit losses. For this reason, it is fundamental to understand and predict the tail distribution of financial returns. As claimed in Rocco (2014), a portfolio is more affected by a few extreme movements in the market than by the sum of many small movements. This motivates risk managers to be primarily concerned with avoiding big unexpected losses. The tool to perform inference over such unexpected events is *extreme value theory* (EVT) which provides a coherent probabilistic framework to model the tail of a distribution. Standard EVT methods require a number of assumptions which are usually hard to verify in practice: usually the choice of a threshold over which observations are assumed to be extreme.

Extreme value mixture models (Scarrott and MacDonald (2012)) have been introduced to provide an objective, data-driven way to estimate the threshold (so overcoming the subjectivity of traditional graphical diagnostics) as well as to account for the uncertainty associated with the threshold choice. Although some non-stationary extensions exist (Lima, Nascimento and Ferraz (2018), Nascimento, Gamerman and Davis (2016), Nascimento, Gamerman and Lopes (2016)), such models are not capable of explicitly taking into account that financial returns are often destabilized by shocks concurring with periods of different extreme behaviours.

---

*Key words and phrases.* Extreme value mixture models, financial returns, GPD distribution, high quantiles, threshold estimation.

Received February 2020; accepted June 2021.

A new class of models is introduced here, termed *change-point extreme value mixture models*, which, whilst estimating in a data-driven fashion of the threshold, are also able to formally represent different extreme regimes caused by financial shocks. This approach not only correctly identifies the location of such shocks, but also gives model-based uncertainty measures about them.

Inference is carried out within the *Bayesian* paradigm using the MCMC machinery (Gamerman and Lopes (2006)), enabling us to easily deliver a wide variety of estimates and predictions of quantities of interest, for example, high quantiles.

Before formally defining our approach, univariate EVT and non-stationary models are reviewed to highlight the relevance and the novelty of our methods.

## 1.1 Extreme value theory

A common approach to model extremes, often referred to as *peaks over threshold*, studies the exceedances over a threshold. A key result to apply this methodology is due to Pickands (1975) which states that, under general regularity conditions, the only possible non-degenerate limiting distribution of properly rescaled exceedances over a threshold  $u$  is the *generalized Pareto distribution* (GPD). Its distribution function (df)  $G$  is defined as

$$G(x|\xi, \sigma, u) = \begin{cases} 1 - \left(1 + \xi \frac{x - u}{\sigma}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{x - u}{\sigma}\right), & \text{if } \xi = 0, \end{cases}$$

for  $u, \xi \in \mathbb{R}$  and  $\sigma \in \mathbb{R}_+$ , where the support is  $x \geq u$  if  $\xi \geq 0$  and  $u \leq x \leq u - \sigma/\xi$  if  $\xi < 0$ . Therefore, the GPD is bounded if  $\xi < 0$  and unbounded from above if  $\xi \geq 0$ . The application of this result in practice entails first the selection of a threshold  $u$  beyond which the GPD approximation appears to be tenable and then the fit of a GPD over data points that exceed the chosen threshold. Thus only a small subset of the data points, those beyond the chosen threshold, are formally retained during the inferential process.

The choice of the threshold over which to fit a GPD is subjective. Although tools to guide this choice exist (Davison and Smith (1990), DuMouchel (1983)), inference can greatly vary for different thresholds (Scarrott and MacDonald (2012), Tancredi, Anderson and O'Hagan (2006)).

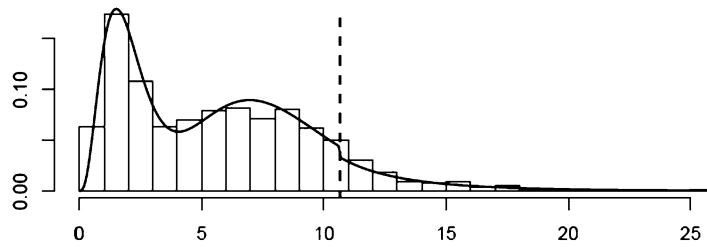
## 1.2 Extreme value mixture models

To overcome the difficulties associated with the selection of a threshold, a variety of models called *extreme value mixture models* (Scarrott and MacDonald (2012)) have been recently defined, which formally use the full dataset and do not require a fixed threshold. These combine a flexible model for the bulk of the data below the threshold, a formally justifiable distribution for the tail and uncertainty measures for the threshold.

The density function  $f$  of an extreme value mixture model is defined as

$$f(x|\Phi, \Psi) = \begin{cases} h(x|\Phi), & x \leq u, \\ [1 - H(u|\Phi)]g(x|\Psi), & x > u, \end{cases} \quad (1.1)$$

where  $h$  is the density, parametrized by  $\Phi$ , of the bulk, that is, the portion of data below the threshold  $u$ ,  $H$  its is df and  $g$  is the GPD density function with parameters  $\Psi = \{\xi, \sigma, u\}$ , which models the tail of the distribution above the threshold  $u$ . An example of an MGPD density fitting simulated data is presented in Figure 1, where it is clearly discernible that the bulk of the distribution consists of a mixture of 2 Gammas, whilst beyond the threshold the



**Figure 1** Example of a MGPD density fit consisting of a mixture of 2 Gammas for the bulk: solid line—MGPD density; dashed line—threshold.

density has GPD decay. Notice that the density  $f$  in equation (1.1) is not necessarily continuous at the threshold: however, Nascimento, Gamerman and Lopes (2012) demonstrated that continuity can be imposed by considering the predictive density.

The first proposal to use the full dataset to estimate both the threshold location and the tail of the distribution is due to Behrens, Lopes and Gamerman (2004), which used a Gamma for the bulk. Since then a variety of proposal for the bulk have been used, including a Normal distribution (Carreau and Bengio (2009)), an infinite mixture of Uniforms (Tancredi, Anderson and O’Hagan (2006)), a mixture of Gammas (Nascimento, Gamerman and Lopes (2012)) and a kernel estimator (MacDonald et al. (2011)). A related approach to estimate tails using the full dataset has been defined in de Carvalho et al. (2020).

Nascimento, Gamerman and Lopes (2012) demonstrated that nothing is lost in extreme estimation by using the full dataset in cases where the determination of the threshold is easy. Conversely, when uncertainty about the threshold location is high, extreme value mixture models outperform the standard peaks over threshold approach in terms of return levels and parameters estimation.

### 1.3 Non-stationary extremes

The above methods assume that all observed data come from a same underlying distribution independently. However, in financial, as well as ecological, applications the structure and amplitude of extremes events usually changes through time. For this reason, inference over financial extremes is often carried out using dynamic models. In this direction, Bollerslev (1987) used a GARCH(1,1) model with Student-T innovations to explicitly take into account of the longer tails often encountered in financial datasets.

Dynamic models based on EVT then started to appear. For instance, McNeil and Frey (2000) proposed a two-stage approach where dependence is first removed using a GARCH model followed by GPD estimation to the assumed independent residual innovations. In a Bayesian setting, Huerta and Sansó (2007) proposed a hierarchical dynamic model based on the generalized extreme value (GEV) distribution, whilst Zhao et al. (2011) defined a GARCH model directly over the GEV parameters. Dynamic extensions of extreme value mixture models have been recently defined (Lima, Nascimento and Ferraz (2018), Nascimento, Gamerman and Lopes (2016)).

Although the above approaches take into account the time dependent nature of rare events, in financial settings extreme variations occur by sudden shocks caused by exogenous agents as described, for instance, by Caldara et al. (2016) and Dierckx and Teugels (2010). Change-point models allow for changes of the model distribution at multiple unknown time points and therefore can be faithfully used to represent and make inference over financial shocks. Some of the first change-point models using the Bayesian MCMC machinery are due to Albert and Chib (1993) and Carlin, Gelfand and Smith (1992), which were extended to multiple change-points in Stephens (1994). Since then the number of change-point models proposed in the

literature has increased dramatically (see, e.g., Horváth and Rice (2014)). However, change-point models which explicitly study distributional changes in the structure of the extremes are very limited.

One of the most common solutions is to define a Markov switching model over the parameters of a GARCH-like model (e.g., Gray (1996), Sampid, Hasim and Dai (2018), Samuel (2008)). However these approaches do not formally use EVT to estimate the tail distribution in each regime. Concerning EVT-based approaches, in the frequentist setting Dierckx and Teugels (2010) and Jarušková and Rencová (2008) defined an hypothesis testing routine to investigate the presence of change-points in GPD and GEV distributions, respectively. In the Bayesian setting, Nascimento and Moura e Silva (2017) developed MCMC algorithms to identify change-points in GEV models. A highly flexible, new approach for inference over extremes in GPD models is proposed here which not only estimates the location of extreme change-points but also the extreme behavior within each regime by using the full dataset.

## 1.4 Outline of the paper

Our approach and inferential routines are next described in Section 2. Section 3 presents a simulation study to both investigate their performance and address the issue of model choice. In Section 4, our methodology is applied to two real-world financial applications: 2-days maxima absolute returns of the NASDAQ stock and negative daily returns of the Royal Bank of Scotland (RBS) stock. A discussion concludes the paper.

## 2 Change-point extreme value mixture models

Let  $x_1, x_2, \dots, x_n$  be a series of time-ordered observations. The probability density function of a change-point extreme value mixture model is defined as

$$f(x_t|\Phi, \Psi, \tau) = \begin{cases} h(x_t|\Phi), & x_t \leq u_j, t \in (\tau_{j-1}, \tau_j], j \in [k], \\ [1 - H(u_j|\Phi)]g_j(x_t|\Psi_j), & x_t > u_j, t \in (\tau_{j-1}, \tau_j], j \in [k], \end{cases} \quad (2.1)$$

where  $h$  is a model parametrized by  $\Phi$  for the bulk below the threshold  $u_j$ ,  $H$  its df,  $g_j$  a GPD density whose parameters are  $\Psi_j = \{u_j, \xi_j, \sigma_j\}$ ,  $\tau = \{\tau_0, \dots, \tau_k\}$  the change-point locations,  $\Psi = \{\Psi_1, \dots, \Psi_k\}$  and  $[k] = \{1, \dots, k\}$ . The parameters of the GPD vary according to the regime in which the observations above the regime-dependent threshold are situated, whilst the bulk distribution  $h$  is common to all regimes and does not vary. The change-points mark a distributional change in the extremes only, and not on the overall distribution of the data. The change-points are integer values corresponding to the index of an observations that mark a sudden change in the distribution of the data. In this setting,  $\tau_0 = 0$  and  $\tau_k = n$ : thus there are  $k - 1$  inner change-points and  $k$  extreme regimes.

Change-point extreme value mixture models have the very useful property of a parametric closed form for expected return levels above the threshold in each regime. The expected return level for each  $t$  period in time is defined as the  $1 - \frac{1}{t}$  quantile, that is, the value  $r_t$  for which an equal or higher value is expected to occur once every  $t$  periods of time. From Nascimento, Gamerman and Lopes (2012), a return  $r_{j,t}$  above the threshold in regime  $j$  is given by

$$r_{j,t} = u_j + \frac{\sigma_j}{\xi_j} ((1 - p_j^*)^{-\xi_j} - 1) \quad \text{where } p_j^* = \frac{1 - \frac{1}{t} - H(u_j|\Phi)}{1 - H(u_j|\Phi)}. \quad (2.2)$$

The model definition in equation (2.1) is general and for practical purposes it needs to be refined by a specific choice of density  $h$ . Two possible choices based on finite mixtures are introduced next, but in general  $h$  can be any density over which Bayesian inference can be carried out.

### 2.1 The CMGPD model

When the common distribution  $h$  for the bulk is a finite mixture of Gammas, the change-point extreme value mixture model is called a CMGPD $_l^k$  model, where  $l$  denotes the number of mixture components and  $k$  the number of different extreme regimes. The CMGPD model extends the MGPD of Nascimento, Gamerman and Lopes (2012) to include extreme change-points. A finite mixture of  $l$  Gammas is defined as  $h(x_t|\Phi) = \sum_{i \in [l]} p_i f_G(x_t|\mu_i, \eta_i)$ , where  $f_G$  is a Gamma density parametrized by the mean  $\mu_i$  and the shape  $\eta_i$ , that is,

$$f_G(x_t|\mu_i, \eta_i) = (\eta_i/\mu_i)^{\eta_i} \Gamma(\eta_i)^{-1} x_t^{\eta_i-1} \exp\{-(\eta_i/\mu_i)x\}, \quad \text{for } x_t > 0,$$

with  $\mu_i, \eta_i \in \mathbb{R}_+$  and  $p_i \in [0, 1]$  such that  $\sum_{i \in [l]} p_i = 1$ . The parametrization in terms of mean and shape parameters is used to solve identifiability issues (Wiper, Rious Insua and Ruggeri (2001)). In this setting  $H(x|\Phi) = \sum_{j \in [l]} p_j F_G(x|\Phi)$ , where  $F_G$  is the Gamma df. The CMGPD model can be used to fit data over the positive real line, as for instance, absolute financial returns.

### 2.2 The CMNPD model

The CMNPD model is similarly defined to the CMGPD, with the difference that the bulk distribution is now a finite mixture of Normal distributions. Formally,  $h(x_t|\Phi) = \sum_{j \in [l]} p_j f_N(x_t|\mu_j, \delta_j^2)$ , where  $f_N(x_t|\mu_j, \delta_j^2)$  is the Normal density with mean  $\mu_j \in \mathbb{R}$  and variance  $\delta_j^2 \in \mathbb{R}_+$ . Thus, this model is used in financial applications where data takes values on the real line and interest is on the right tail only.

### 2.3 Prior distribution

The model definition is completed by prior distributions for the parameters. GPD parameters of different regimes are a priori assumed independent. The prior distribution for  $(\xi_j, \sigma_j)$ ,  $j \in [k]$ , is the non-informative prior of (Castellanos and Cabras, 2007) defined as  $\pi(\xi_j, \sigma_j) \propto \sigma_j^{-1} (1 + \xi_j)^{-1} (1 + 2\xi_j)^{-1/2}$ .

The priors for the different regimes' thresholds are independent Normal, distributions as suggested by Behrens, Lopes and Gamerman (2004), with mean  $\mu_u$  chosen around a high order sample statistic, since the GPD approximation can be expected to hold only over the tail of the data. The variance  $\sigma_u^2$  is chosen so that the bulk of the prior distribution ranges roughly over data points in the upper half. This is the only partially informative prior used, but the hyper-parameter choices are guided by the theoretical results of Pickands (1975). Thus,  $\pi(u_j) = f_N(\mu_u, \sigma_u^2)$ .

The change-points are given the discrete uniform distribution subject to the restriction  $\{\tau_0 < \tau_1 < \dots < \tau_k\}$ , as suggested by Stephens (1994):

$$\pi(\tau_1, \dots, \tau_k) = \frac{1}{\tau_2} \mathbb{1}_{(1 \leq \tau_1 < \tau_2)} \frac{1}{\tau_3 - \tau_1} \mathbb{1}_{(\tau_1 < \tau_2 < \tau_3)} \dots \frac{1}{n - \tau_{k-2}} \mathbb{1}_{(\tau_{k-2} < \tau_{k-1} \leq n)}.$$

The prior distribution for the bulk density parameter  $\Phi$  depends on the model used. In both cases, the weights  $(p_1, \dots, p_l)$  are given a Dirichlet prior with parameter  $(1, \dots, 1)$ . For the CMGPD model, the prior definition is as in Nascimento, Gamerman and Lopes (2012). Each shape parameter  $\eta_j$  is given an independent Gamma prior  $\pi(\eta_j) = f_G(\eta_j|c_j/d_j, c_j)$ , where  $c_j, d_j \in \mathbb{R}_+$  are chosen to give a large prior variance. The prior for the Gamma means is  $\pi(\mu_1, \dots, \mu_l) = K \prod_{j \in [l]} f_{IG}(\mu_j|a_j, b_j) \mathbb{1}_{(0 < \mu_1 < \dots < \mu_l)}$ , where  $f_{IG}$  is the inverse Gamma density,  $K$  is a normalizing constant and  $a_j$  and  $b_j$  are chosen to give a large prior variance. The order restriction over the means is set to ensure identifiability.

For the CMNPD model, priors for the Normal mixture parameters  $(\mu_j, \delta_j^2)_{j \in [l]}$  are given as follows. The prior for the means is given conditionally on the variances as

$\pi(\mu_1, \dots, \mu_l | \delta_1, \dots, \delta_l) = \prod_{j \in [l]} f_N(\mu_j | j/M, (\alpha/\delta_j)^2) \mathbb{1}_{(\mu_1 < \dots < \mu_l)}$ , where  $\alpha$  is chosen to give a large prior variance and  $M = \max(x_1, \dots, x_n)$ . This choice is motivated by the symmetry of financial returns, so to assure the closeness of the means to 0, and by identifiability issues. Each mixture variance is independently given a Gamma distribution, that is,  $\pi(\delta_j^2) = f_G(\delta_j^2 | c_j/d_j, c_j)$  where again hyperparameters are chosen to give large variances.

The overall prior for a change-point extreme value mixture model can be written as  $\pi(\Phi, \Psi, \tau) = \pi(\Phi)\pi(\tau) \prod_{j \in [k]} \pi(\xi_j, \sigma_j)\pi(\mu_j)$ .

## 2.4 Posterior inference

For a sample  $x = (x_1, \dots, x_n)$  the log-posterior of the CMGPD $_l^k$  model,  $\log \pi(\Phi, \Psi, \tau | x)$ , up to an additive constant is given by

$$\begin{aligned} & \sum_{j \in [k]} \sum_{t: x_t \leq u_j} \log \left( \sum_{z \in [l]} p_z f_G(x_t | \mu_z, \eta_z) \right) \mathbb{1}_{(t \in (\tau_{j-1}, \tau_j])} \\ & + \sum_{j \in [k]} \sum_{t: x_t > u_j} \log \left( 1 - \sum_{z \in [h]} p_z F_G(u_j | \mu_z, \eta_z) \right) \mathbb{1}_{(t \in (\tau_{j-1}, \tau_j])} \\ & + \sum_{j \in [k]} \sum_{t: x_t > u_j} \log(g(x_t | \Psi_j)) \mathbb{1}_{(t \in (\tau_{j-1}, \tau_j])} + \log(\pi(\Phi, \Psi, \tau)) \end{aligned} \quad (2.3)$$

For the CMNPD $_l^k$  model the log-posterior can be easily deduced by substituting  $f_G$  and  $F_G$  in equation (2.3) with  $f_N$  and  $F_N$  respectively.

Inference cannot be performed analytically and approximating MCMC algorithms are used. Parameters are divided into blocks and updating of the blocks follows Metropolis-Hastings steps since full conditionals have no recognizable form. Proposal variances are tuned via an adaptive algorithm as suggested in Roberts and Rosenthal (2009). Details are given in the Supplementary Material. All algorithms are implemented in R.

Summaries of financial extreme returns can be straightforwardly computed from the posterior distribution. Common measures used for financial losses are the Value-at-Risk (VaR) and the expected shortfall (ES). VaR is generally defined as the risk capital sufficient to cover losses from a portfolio over a holding period of a fixed number of days. It corresponds to the  $p^{th}$  quantile over a certain time horizon and is denoted as  $\text{VaR}_p$ . ES, or tail conditional expectation, is defined as the potential size of a loss exceeding a specific  $\text{VaR}_p$ . It corresponds to the expectation conditional on observing values larger than  $\text{VaR}_p$ . For change-point extreme value mixture models, the expected shortfall in the  $j$ -th regime takes the closed form

$$\text{ES}_p = \frac{\text{VaR}_{p,j}^t}{1 - \xi_j} + \frac{\sigma_j - \xi_j u_j}{1 - \xi_j}, \quad (2.4)$$

where  $\text{VaR}_{p,j}$  is the Value-at-Risk in regime  $j$ .

Both VaR and ES are highly non-linear functions of the model's parameters (see equations (2.2) and (2.4)). Thus, their posterior distribution cannot be derived analytically. However, the MCMC machinery enables us to derive an approximated distribution for *any* function of the models' parameters, as demonstrated in our applications in Section 4.

## 3 Simulation study

A simulation study is conducted next with two main purposes: first, to assess the identifiability of the models proposed; second, to validate model selection criteria. For brevity the results

for data generated from one specific instance of CMGPD and MGPD are reported here. Additional results are reported in Appendix B. Two samples of 5000 observations were generated, one from a MGPD<sub>2</sub>, the other from a CMGPD<sub>2</sub><sup>3</sup>, where the subscript denotes the number of mixture components and the superscript the number of extreme regimes. In both datasets the mixture parameters are  $(\mu_1, \mu_2) = (2, 8)$ ,  $(\eta_1, \eta_2) = (4, 8)$  and  $(p_1, p_2) = (2/3, 1/3)$ . For the MGPD data, GPD parameters were fixed at  $\xi = 0.4$  and  $\sigma = 2$ , whilst the threshold was placed at the 85th theoretical quantile of the Gamma mixture (7.99).

The simulated observations from CMGPD<sub>2</sub><sup>3</sup> had change-point locations  $\tau = \{2000, 3500\}$ . The regime-dependent GPDs were chosen so that  $(\xi_1, \xi_2, \xi_3) = (-0.4, 0, 0.4)$ ,  $(\sigma_1, \sigma_2, \sigma_3) = (0.5, 1, 1.5)$  and the regimes' thresholds were placed respectively, at the 80th (6.99), 85th (7.99) and 90th (9.22) theoretical quantiles.

Simulations were run on a PC with processor 2,7 GHz Intel Core i5 and 8 Gb RAM. For all simulations, the codes ran for 15,000 iterations, with a burn-in of 5000 and thinning every 10, giving a posterior sample of 1000. Convergence was assessed by running parallel chains with different starting values and then comparing the resulting estimates.

In all cases, to reduce the number of models to be compared, the number of mixture components was first chosen by fitting MGPD<sub>*l*</sub> and CMGPD<sub>*l*</sub><sup>3</sup> for various *l*. As already shown in Nascimento, Gamerman and Lopes (2012) and Leonelli and Gamerman (2020), the correct number of mixture components can be retrieved from the posterior sample since the weights of all extra components are estimated as zero. Table 1 further shows that adding extra mixture components does not improve the model fit.

Having fixed the number of mixtures, models with varying change-points' numbers were fitted to the simulated datasets. As already discussed in Leonelli and Gamerman (2020), standard model selection criteria often fail to identify the correct model in the setting of extreme value mixture models. This is shown in Table 1 where BIC (Schwarz (1978)) and DIC (Spiegelhalter et al. (2002)) fail to select the true model. Conversely, the true generating model is always preferred by the WAIC of Watanabe (2010). This criterion has been shown to be particularly robust for mixtures and non-identifiable models.

The number of regimes can be further identified when fitting CMGPD with non-necessary change-points since the exceeding locations converge to values very close to 0, *n* or another change-point, depending on the starting values of the MCMC algorithm, as noted in Nascimento and Moura e Silva (2017). This can be seen in Table 2. When a CMGPD<sub>2</sub><sup>4</sup> model is estimated over CMGPD<sub>2</sub><sup>3</sup> data, the two change-points are correctly identified, whilst the third is located close to zero. Conversely, fitting a CMGPD<sub>2</sub><sup>2</sup> model over CMGPD<sub>2</sub><sup>3</sup> data the only change-point is estimated around the true change-point giving a larger distributional

**Table 1** Model selection criteria for models estimated over simulated datasets

Data	Model	BIC	DIC	WAIC
CMGPD <sub>2</sub> <sup>3</sup>	MGPD <sub>2</sub>	21,212.55	21,148.29	22,037.79
CMGPD <sub>2</sub> <sup>3</sup>	MGPD <sub>3</sub>	21,246.17	21,155.14	22,044.63
CMGPD <sub>2</sub> <sup>3</sup>	MGPD <sub>4</sub>	21,266.49	21,148.84	22,039.89
CMGPD <sub>2</sub> <sup>3</sup>	CMGPD <sub>2</sub> <sup>2</sup>	20,414.07	20,361.69	20,366.07
CMGPD <sub>2</sub> <sup>3</sup>	CMGPD <sub>2</sub> <sup>3</sup>	<b>20,309.73</b>	20,255.33	<b>20,262.07</b>
CMGPD <sub>2</sub> <sup>3</sup>	CMGPD <sub>2</sub> <sup>4</sup>	20,320.46	<b>20,255.15</b>	20,263.74
MGPD <sub>2</sub>	MGPD <sub>2</sub>	22,654.83	22,592.46	<b>22,595.90</b>
MGPD <sub>2</sub>	MGPD <sub>3</sub>	22,686.81	22,588.36	22,597.55
MGPD <sub>2</sub>	MGPD <sub>4</sub>	22,744.25	22,559.29	22,037.79
MGPD <sub>2</sub>	CMGPD <sub>2</sub> <sup>2</sup>	<b>22,649.49</b>	<b>22,574.54</b>	22,597.96

**Table 2** Posterior distribution of change-point locations for CMGPD<sub>2</sub><sup>2</sup> (left), CMGPD<sub>2</sub><sup>3</sup> (centre) and CMGPD<sub>2</sub><sup>4</sup> (right). True values are  $\tau_1 = 2000$  and  $\tau_2 = 3500$ 

	$\tau_1$	$\tau_2$	$\tau_3$
CMGPD <sub>2</sub> <sup>2</sup>	2000 (1996,2002)		
CMGPD <sub>2</sub> <sup>3</sup>	2000 (1996,2002)	3464 (3332,3562)	
CMGPD <sub>2</sub> <sup>4</sup>	7 (1,31)	2000 (1996,2002)	3465 (3323,3566)

**Table 3** Posterior means and 95% credibility intervals for the parameters of simulated CMGPD<sub>2</sub><sup>3</sup> data related to the tail densities ( $\xi$ ,  $\sigma$  and  $u$ )

Parameter	$\xi_1 = -0.4$	$\xi_2 = 0$	$\xi_3 = 0.4$
MGPD <sub>2</sub>	0.30 (0.23,0.39)		
CMGPD <sub>2</sub> <sup>3</sup>	-0.38 (-0.43, -0.31)	0.01 (-0.11, 0.18)	0.49 (0.29, 0.76)
Parameter	$\sigma_1 = 0.5$	$\sigma_2 = 1$	$\sigma_3 = 1.5$
MGPD <sub>2</sub>	1.09 (0.99, 1.21)		
CMGPD <sub>2</sub> <sup>3</sup>	0.48 (0.43, 0.53)	0.98 (0.80, 1.19)	1.52 (1.12, 1.96)
Parameter	$u_1 = 6.99$	$u_2 = 7.99$	$u_3 = 9.22$
MGPD <sub>2</sub>	6.99 (6.99, 7.00)		
CMGPD <sub>2</sub> <sup>3</sup>	6.99 (6.99, 7.00)	8.02 (8.00, 8.04)	9.07 (8.80, 9.25)

change: in this case the one at  $t = 2000$  associated to a switch from an upper bounded distribution to an unbounded one. The histograms further show that in all cases, uncertainty about the strongest change-point is limited, whilst the posterior distribution for the change-point located at  $t = 3500$  has a larger variance. The same conclusion can be drawn when fitting a CMGPD<sub>2</sub><sup>2</sup> model over MGPD<sub>2</sub> data, since the posterior mean of the only change-point is 40.25 with 95% credibility interval (9, 88).

Having ensured that the true model can be correctly chosen, the identifiability of the parameters is investigated next. As in Nascimento, Gamerman and Lopes (2012), it can be shown that all bulk parameters are correctly estimated. But more interestingly, Table 3 demonstrates that tail parameters are well estimated for all regimes when using data simulated from the CMGPD<sub>2</sub><sup>3</sup> model. When an MGPD<sub>2</sub> model is fitted over this dataset, each tail parameter is estimated around an average value of those of all regimes. When the CMGPD<sub>2</sub><sup>2</sup> model is fitted over MGPD<sub>2</sub> data, the parameters associated to the non-empty regime well estimate the true tail parameters, as shown in Table 4.

Given the use of vague priors, there is a clear indication that the likelihood can correctly identify the true values. In particular, the estimation of the tail parameters is highly successful evidencing the ability of the model to recover varying tail behavior.

## 4 Applications

### 4.1 NASDAQ absolute daily returns

The first financial dataset considered consists of absolute daily returns of NASDAQ stock values from January 1996 to December 2017. In order to exclude returns equal to zero, maxima of sets of 2 days are taken for a total of 2768 observations (similar pre-processing steps are taken, for instance, in Nascimento, Gamerman and Lopes (2012)). The aim is the estimation of volatility of the composite index over time comparing the MGPD and the CMGPD approaches.

**Table 4** Posterior means and 95% credibility intervals for the parameters of simulated MGPD<sub>2</sub> data with true parameters  $\xi = 0.4$ ,  $\sigma = 2$  and  $u = 8.02$ 

	$\xi_1$	$\xi_2$	$\sigma_1$	$\sigma_2$	$u_1$	$u_2$
MGPD <sub>2</sub>	0.43 (0.33, 0.55)		2.02 (1.76, 2.46)		8.28 (7.82, 9.44)	
CMGPD <sub>2</sub>	0.09 (-0.50, 0.59)	0.45 (0.35, 0.59)	1.72 (0.15, 4.01)	2.04 (1.77, 2.52)	9.07 (7.37, 12.93)	8.39 (8.00, 9.37)

**Table 5** Model selection criteria for models estimated over NASDAQ and RBS data

NASDAQ				RBS			
Model	BIC	DIC	WAIC	Model	BIC	DIC	WAIC
MGPD <sub>1</sub>	7527.67	7488.93	7329.53	MNPD <sub>1</sub>	22,425.55	22,422.27	22,289.93
MGPD <sub>2</sub>	7543.31	7486.00	7488.20	MNPD <sub>2</sub>	21,761.30	21,735.96	21,599.13
MGPD <sub>3</sub>	7570.16	7480.38	7485.45	MNPD <sub>3</sub>	21,716.40	21,669.94	21,526.55
MGPD <sub>4</sub>	7603.38	7472.06	7486.89	MNPD <sub>4</sub>	21,757.50	21,683.06	21,538.88
CMGPD <sub>1</sub> <sup>5</sup>	7492.97	7408.87	7203.87	CMNPD <sub>1</sub> <sup>5</sup>	21,510.00	21,413.04	21,462.25
CMGPD <sub>1</sub> <sup>6</sup>	<b>6696.19</b>	6615.71	<b>6776.51</b>	CMNPD <sub>1</sub> <sup>6</sup>	<b>21,424.62</b>	21,356.38	<b>21,381.23</b>
CMGPD <sub>1</sub> <sup>7</sup>	6719.61	<b>6612.89</b>	7106.92	CMNPD <sub>1</sub> <sup>7</sup>	21,429.53	<b>21,345.34</b>	21,392.57
CMGPD <sub>2</sub> <sup>6</sup>	7080.70	6980.42	7022.76	CMNPD <sub>2</sub> <sup>6</sup>	22,052.55	22,022.27	22,078.54

The number of Gamma components for the bulk is first investigated and it is observed that only one component is needed, with estimated parameters in the selected model  $\mu = 3.3$  (2.8, 3.8) and  $\eta = 1.3$  (1.2, 1.5). The number of change-points is chosen resorting to information criteria and posterior locations. The most reliable WAIC criterion favours a model with 6 regimes as reported in Table 5, which also includes evidence that a change-point approach outperforms the static one. This is confirmed in Table 6 reporting the posterior distribution of the change-points: the posterior mean of the first CMGPD<sub>1</sub><sup>7</sup> change-point equals 4, thus giving an empty regime and confirming the optimality of CMGPD<sub>1</sub><sup>6</sup>.

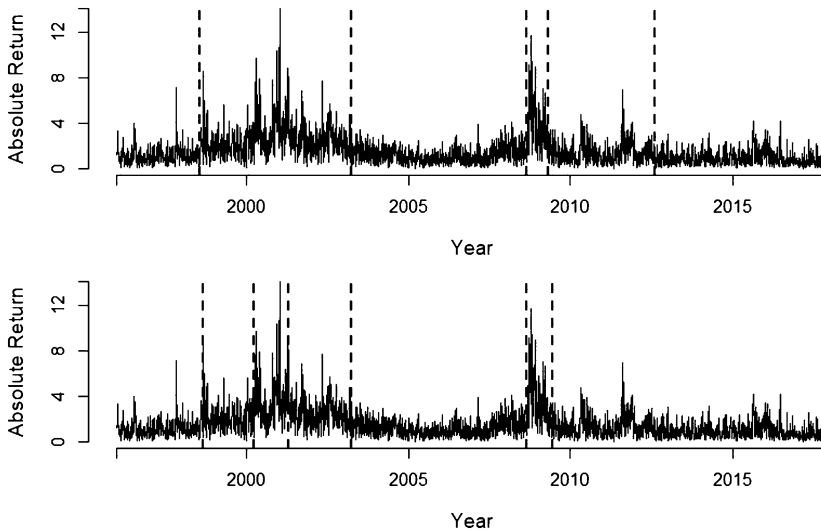
The posterior means of change-points from the CMGPD<sub>1</sub><sup>6</sup> model are located on July 1998, April 2003, August 2008, May 2009 and April 2012 as shown in Figure 2 top. An alternation of regimes with low/medium volatility and high volatility can be noticed, so different tail parameters and returns are to be expected. To assess the validity of the CMGPD estimation, change-points are also estimated using the `changepoint` R package (Killick and Eckley (2014)). In details, the function `cpt.var` is used with estimation method `PELT`, penalty `BIC` and a minimum regime length of 100 observations. The results from the two procedures overall agree as shown in Figure 2: the `changepoint` package splits the second regime into three smaller regimes and merges the last two regimes in a unique one.

Table 7 summarizes the posterior distribution of the CMGPD<sub>1</sub><sup>6</sup> tail parameters. This demonstrates the flexibility of our approach of discriminating between periods of high and low volatility: in the 2nd and 4th regimes the estimates of the scale  $\sigma$  and shape  $\xi$  parameters are larger than all other regimes demonstrating higher level of stress of the market. The values of the estimated thresholds suggest a particular behaviour of this dataset: the 1st, 3rd, 5th, 6th regimes resemble more a GPD distribution than a MGPD. As a result, the estimated thresholds for these regimes are very close to 0 and the proportion of observations below the threshold is small. The flexibility of the model proposed enable us to take this into account without any complication. Furthermore, it has already been observed in other extreme value mixture models that the threshold may be estimated at a relatively small value (see, e.g., Lima, Nascimento and Ferraz (2018), Nascimento, Gamerman and Lopes (2011)). Conversely, one could set up a prior distribution for the thresholds more concentrated around higher empirical quantiles of the dataset to ensure that a bigger proportion of data points fall below the threshold (see Nascimento, Gamerman and Lopes (2012), for details).

The expected return levels, for  $t \in [10, 1000]$ , are reported in Figure 3 for each estimated regime. The CMGPD estimates and their 95% confidence interval (shaded area) are fairly close to the empirical ones in all regimes, thus confirming the goodness of our estimation routines. Returns were further estimated using the MGPD and the GPD (using the threshold estimated by the CMGPD). The CMGPD estimates are always closer to the empirical values

**Table 6** *Posterior distribution of change-point locations for models estimated over NASDAQ data*

	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$	$\tau_6$
CMGPD <sub>1</sub> <sup>5</sup>	918 (702, 1036)	1336 (921, 1599)	1602 (1581, 1636)	1645 (1626, 1673)		
CMGPD <sub>1</sub> <sup>6</sup>	323 (318, 327)	915 (907, 929)	1594 (1588, 1598)	1681 (1668, 1695)	2049 (2008, 2094)	
CMGPD <sub>1</sub> <sup>7</sup>	4 (1, 8)	323 (318, 326)	915 (907, 926)	1595 (1590, 1598)	1679 (1667, 1690)	2029 (2006, 2092)



**Figure 2** *NASDAQ 2-day max absolute returns time series with estimated change-points using CMGPD<sub>1</sub><sup>6</sup> (top) and the changepoint R package (bottom).*

than the MGPD ones. Furthermore, the GPD estimates overlays the CMGPD ones in all the regimes with a low threshold, whilst in the others CMGPD clearly outperforms GPD. Thus the use of the full dataset, divided into extreme regimes, leads to better posterior estimates.

## 4.2 Royal bank of Scotland daily returns

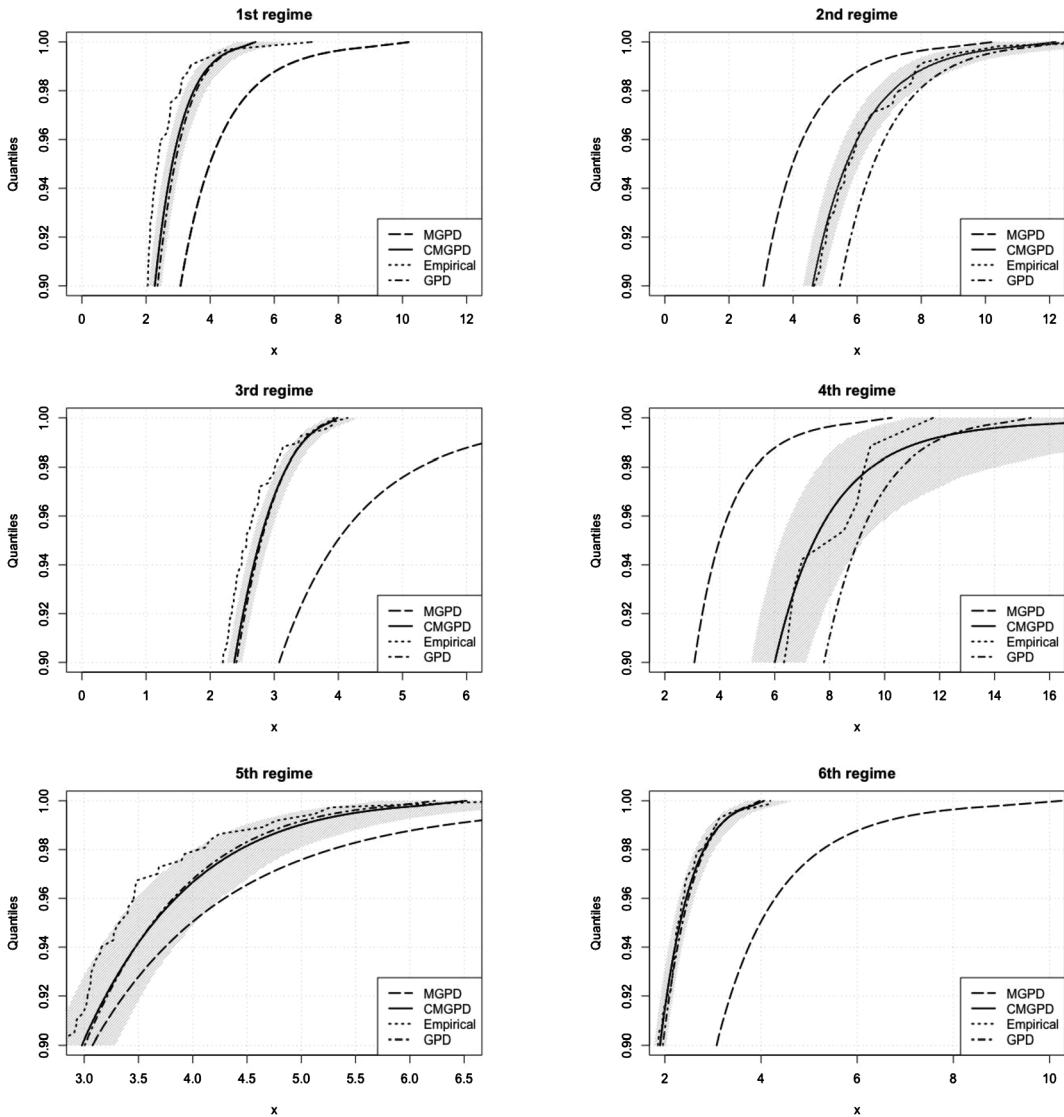
The second financial dataset considered is the Royal Bank of Scotland (RBS) stock daily returns from January 2000 to February 2018 for a total of 4635 observations. In this case, positive and negative returns are modeled at the same time and the focus is on the estimation of the lower tail (a change of sign is applied for convenience). The estimation efficiency of MNPD and CMNPD models are now investigated. For all models, it was first observed that only one Normal component is needed, with estimated parameters in the selected model  $\mu = 0.7$  (0.6, 0.8) and  $\sigma^2 = 11.4$  (10.8, 12.2).

Again all model selection criteria favour a change-point approach compared to the static one, as reported in Table 5 and the WAIC chooses a model with 6 regimes. This is also confirmed by the posterior distribution of the first CMNPD<sub>1</sub><sup>7</sup> change-point, located at the beginning of the series with a posterior mean of 61 and 95% credibility interval (36, 78).

The regimes estimated from the CMNPD<sub>1</sub><sup>6</sup> model are reported in Figure 4 top. The estimated change-point are located on April 2003, July 2007, June 2010, June 2011 and August 2012 with posterior distributions shown in Figure 5. The regimes show different magnitude in losses, with tail parameters reported in Table 8. The first and last three regimes represent periods of medium-sized losses, whilst the second one represents a period of high stability. The third regime is by far the most interesting: it is concurrent to the UK's biggest bank failure in history culminated to the Blue Monday Crash in January 2009. The bank eventually managed to survive thanks to the UK bank rescue package issued by the Government. This is the only regime with a clear heavy tail behaviour ( $\xi > 0$ ). The value of  $\sigma$  is constant among the regimes, with the exception of the second regime whose value of  $\sigma$  indicates lower volatility. Although some estimated values of the threshold  $u$  are negative, around half of the datapoints in the associated regimes are below the threshold. As for the Nasdaq application, our modelling approach has the flexibility of modelling situations where a big portion of the data can be fitted by a GPD distribution.

**Table 7** Posterior distributions of  $\xi$ ,  $\sigma$ ,  $u$  and estimated proportion of points below the threshold for  $\text{CMGPD}_1^6$  over NASDAQ data

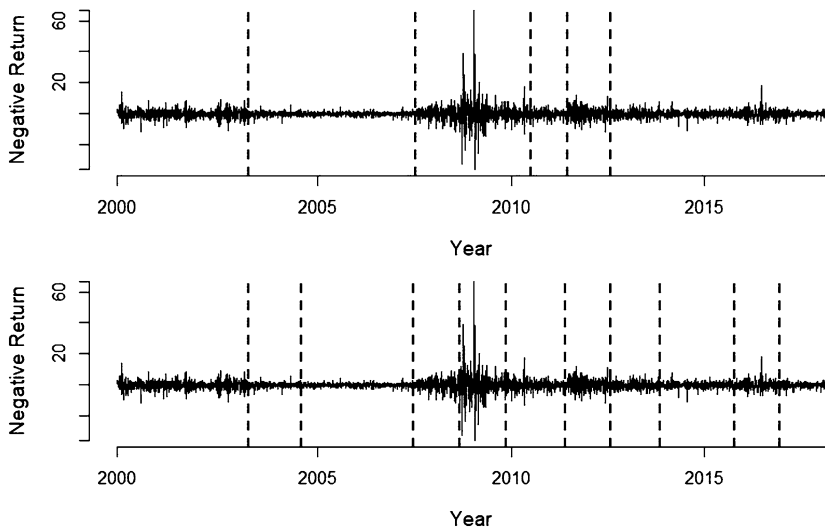
$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$\xi_5$	$\xi_6$
-0.08 (-0.12, -0.08)	0.04 (-0.06, 0.19)	-0.30 (-0.35, -0.22)	0.04 (-0.3, 0.63)	-0.12 (-0.20, -0.002)	-0.14 (-0.20, -0.06)
$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$	$\sigma_6$
0.95 (0.83, 1.08)	1.42 (1.12, 1.66)	1.25 (1.09, 1.42)	2.00 (1.04, 3.28)	1.32 (1.06, 1.59)	0.86 (0.78, 0.97)
$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$
0.30 (0.27, 0.39)	2.00 (1.82, 2.44)	0.32 (0.22, 0.43)	3.31 (2.28, 4.05)	0.44 (0.24, 0.65)	0.22 (0.20, 0.23)
$P(X < u_1)$	$P(X < u_2)$	$P(X < u_3)$	$P(X < u_4)$	$P(X < u_5)$	$P(X < u_6)$
0.06	0.37	0.08	0.47	0.14	0.07



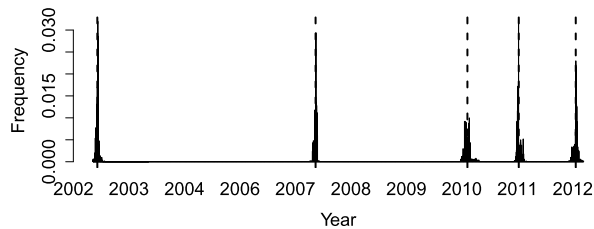
**Figure 3** Return level plots in each estimated regime for NASDAQ data for a sequence of  $t \in [10, 1000]$ , where  $x$  corresponds to the  $1 - 1/t$  quantile.

Again our change-points are compared to those from the `changepoint` package (in this case we used a minimum regime length of 300 observations). Figure 4 shows that whilst `changepoint` selects the same change-points as the CMNPD, it also includes additional ones for a total of 10 change-points. Of course, the output of `changepoint` could include a smaller number of change-points by using a stronger penalization.

Figure 6 reports the crucial VaR estimates from 5% to 0.1% for each regime. The same conclusions can be drawn as in the NASDAQ case, with the CMNPD outperforming both the MNPd and GPD approaches. Table 9 further summarizes our estimates of the expected shortfall at both 5% and 1%. These are compared with the so-called NormFit approach: as reported in Chang et al. (2011) and Gilli and K ellezi (2006) the Basel accord formalizes that financial firms estimate VaR using a Normal hypothesis, which is then multiplied by a ‘safety factor’ of 3 to take account of tail’s heaviness. Whilst NormFit estimates are comparable to ours at the 5% level, they highly underestimate risk at the 1% level.



**Figure 4** RBS daily negative returns time series with estimated change-points using  $\text{CMNPD}_1^6$  (top) and the changepoint R package (bottom).



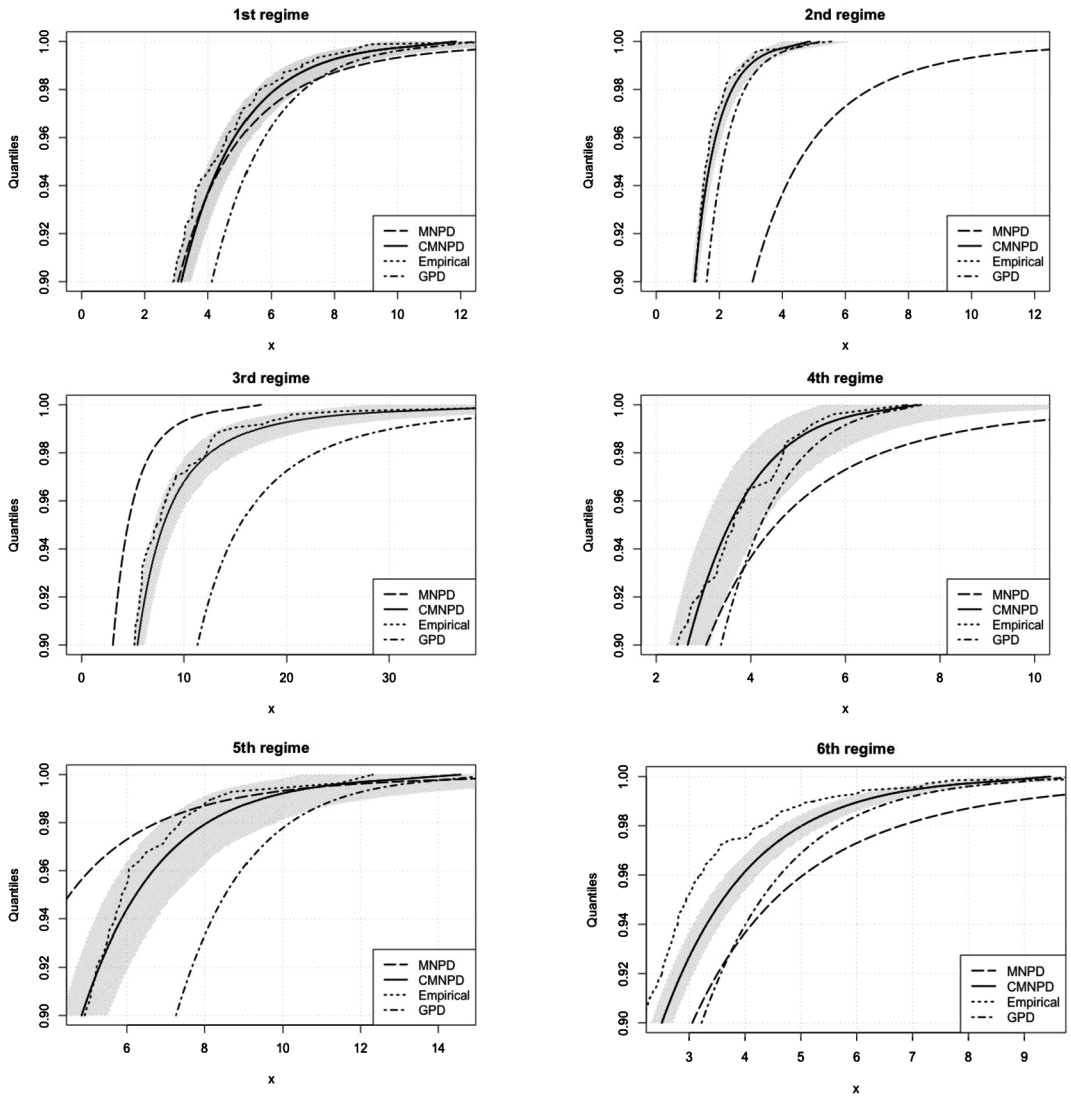
**Figure 5** Posterior histograms of change-point locations for  $\text{CMNPD}_1^6$  estimated over RBS data. Dashed line denote the posterior means.

The performance of the CMNPD model is tested against the GARCH-EVT approach of McNeil and Frey (2000) via *backtesting*: the actual losses at time  $t + 1$  are compared to the estimated VaR at time  $t$ . The backtesting is based on a moving window such that at each time  $t$ , a new set of GARCH(1,1) parameters, residuals and GPD based quantile are estimated. Table 10 reports the number of expected  $\text{VaR}_p$  violations in each regime, equal to  $n(1 - p)$  with  $n$  the number of observations in a regime, and the violations observed from CMNPD and GARCH-EVT. The CMNPD model always outperforms GARCH-EVT in estimating violations for high-volatility regimes (e.g., the 2nd and 3rd) and in very-high quantiles scenarios ( $\text{VaR}_{0.5\%}$  and  $\text{VaR}_{0.1\%}$ ). Thus, CMNPD better estimates the occurrence of very rare events than the GARCH-EVT approach.

The backtesting procedure offers a way to observe the goodness of our VaR estimation but it does not represent a robust model selection criterium. For this reason, the model confidence set procedure devised by Hansen, Lunde and Nason (2011) is utilized here to assess if the CMNPD model is superior to the others. In a nutshell, it is a multiple testing procedure which selects a set of superior models, out of a set of candidate ones, having equal predictive ability for a specific value of VaR. The results of the procedure for the RBS data are summarized in Table 11. The MCS package of Catania and Bernardi (2017) is used with the asymmetric loss function of González-Rivera, Lee and Mishra (2004) and both test statistics,  $T_{\max}$  and  $T_R$ , available in the package to predict  $\text{VaR}_{0.95}$ ,  $\text{VaR}_{0.99}$ ,  $\text{VaR}_{0.995}$  and  $\text{VaR}_{0.999}$ . It can be seen that the  $\text{CMNPD}_1^6$  model not only is the only one always included in the confidence set, but also has the highest value for the test statistic. It can be further noticed that overall models

**Table 8** *Posterior distributions of  $\xi$ ,  $\sigma$ ,  $u$  and estimated proportion of points below the threshold for  $\text{CMNPD}_1^6$  over RBS data*

$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$\xi_5$	$\xi_6$
0.01 (−0.08, 0.10)	0.04 (−0.04, 0.13)	0.37 (0.16, 0.61)	−0.12 (−0.31, 0.08)	−0.01 (−0.25, 0.32)	−0.02 (−0.06, 0.03)
$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$	$\sigma_6$
1.79 (1.57, 2.03)	0.67 (0.58, 0.76)	2.36 (1.79, 3.10)	1.63 (1.23, 2.09)	2.07 (1.37, 3.09)	1.65 (1.52, 1.78)
$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$
−0.008 (−0.02, 0.0002)	0.000 (−0.005, 0.0006)	3.12 (2.23, 3.82)	−0.14 (−0.15, 0.32)	3.00 (2.31, 3.6)	−0.46 (−0.48, −0.43)
$P(X < u_1)$	$P(X < u_2)$	$P(X < u_3)$	$P(X < u_4)$	$P(X < u_5)$	$P(X < u_6)$
0.47	0.44	0.78	0.46	0.75	0.39



**Figure 6** Estimates of  $\text{VaR}_p$ , that is, the  $p$ -quantiles  $x$ , in each estimated regime for values of  $p$  ranging from 0.9 to 0.99 for the RBS data.

based on change-points (namely CMNPD, GPD and NORMIFT), which were identified using our approach, perform better than non-change-point models (MNPD and GARCH-EVT).

### 5 Conclusions

The financial literature asserts that not only extreme behaviour may change considerably in time, but also that these variations occur by sudden shocks which deeply affect volatility scenarios. This work puts forward a change-point generalisation of extreme value mixture models with the ability of detecting multiple change-points in the tail distribution. The inclusion of regime-dependent GPD parameters enables the switch between light and heavy-tailed behaviour explaining well periods of financial stress and market instability.

Due to the semiparametric nature of the models proposed, Bayesian methods are used. Despite the use of vague prior information, our inferential routines recover the correct parameter values, whilst giving uncertainty measures about the crucial threshold and change-point pa-

**Table 9** *ES at 5% and 1% estimated using CMNPD<sub>1</sub><sup>6</sup> and the NormFit approach*

Shortfall	Approach	1st regime	2nd regime	3rd regime	4th regime	5th regime	6th regime
ES <sub>5%</sub>	Empirical	5.87	2.30	13.15	4.63	7.40	4.50
	NormFit	5.16	2.11	11.80	4.40	7.39	4.03
	CMNPD <sub>1</sub> <sup>6</sup>	6.29 (5.56, 7.19)	2.49 (2.23, 2.83)	14.97 (11.50, 21.30)	4.65 (3.85, 5.99)	8.29 (6.98, 10.61)	5.13 (4.73, 5.55)
ES <sub>1%</sub>	Empirical	8.40	3.58	25.45	5.86	10.28	7.52
	NormFit	6.68	2.74	15.20	5.15	9.01	5.22
	CMNPD <sub>1</sub> <sup>6</sup>	9.34 (7.85, 11.45)	3.78 (3.22, 4.56)	30.93 (19.80, 55.32)	6.30 (4.89, 9.03)	10.71 (9.10, 18.80)	7.55 (6.86, 8.37)

**Table 10** Comparing VaR violations with expected violations using CMNPD<sub>1</sub><sup>6</sup> and GARCH-EVT

Value-at-Risk	Approach	1st regime	2nd regime	3rd regime	4th regime	5th regime	6th regime
VaR <sub>5%</sub>	Expected	42	54	37	13	14	70
	CMNPD	40	<b>43</b>	<b>32</b>	<b>14</b>	10	38
	GARCH-EVT	<b>41</b>	40	54	7	<b>15</b>	<b>65</b>
VaR <sub>1%</sub>	Expected	8	10	7	2	3	14
	CMNPD	6	<b>9</b>	<b>7</b>	2	2	8
	GARCH-EVT	<b>9</b>	13	16	2	2	<b>19</b>
VaR <sub>0.5%</sub>	Expected	4	5	3	1	1	7
	CMNPD	<b>3</b>	<b>4</b>	<b>3</b>	<b>1</b>	<b>1</b>	<b>7</b>
	GARCH-EVT	8	8	12	2	2	15
VaR <sub>0.1%</sub>	Expected	1	1	1	0	0	1
	CMNPD	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>2</b>
	GARCH-EVT	4	3	6	1	1	6

**Table 11** Model confidence set for the RBS data and various values of VaR using both  $T_{max}$  and  $T_R$ . Models in the confidence are ranked by the value of the test statistic whilst models outside the set are denote by \*

Model	VaR <sub>0.95</sub>		VaR <sub>0.9</sub>		VaR <sub>0.995</sub>		VaR <sub>0.999</sub>	
	$T_{max}$	$T_R$	$T_{max}$	$T_R$	$T_{max}$	$T_R$	$T_{max}$	$T_R$
CMNPD <sub>1</sub> <sup>6</sup>	1	1	1	1	1	1	1	1
GPD	*	*	2	2	2	2	3	3
NORMFIT	*	*	3	3	4	4	4	4
MNPD	*	*	*	4	3	3	2	2
GARCH-EVT	*	*	*	*	*	*	5	*

rameters. Model choice is easily performed due to the inherent ability of the model to detect the number of both mixture components for the bulk and, most importantly, change-points.

Our approach outperforms all the static and dynamic methods considered for comparison. Since financial markets are heavily affected by unexpected and abrupt variations, extreme regimes are well-captured using change-point tools, identifying periods of changing volatility. Return levels, VaR and ES measures are well estimated by our approach, making it a very powerful tool in a real-data context. Their effectiveness in other fields, for instance environmental and medical applications, is yet to be explored.

Although the number of change-points is correctly identified by model selection criteria, models with a different number of change-points need to be fitted. Approaches to estimate  $k$ , the number of change-points, within our MCMC routines are currently being explored. Recent proposals use the hidden change-point representation of Chib (1998) coupled with a Dirichlet process (e.g., Ko, Chong and Ghosh (2015)). In a nutshell, such models put a uniform prior on the number changepoints, where the maximum number of changepoints is equal to the total number of observations. However, these fail in our context because the acceptance of a new change-point location is based on a subset of the observations. Because our change-points discriminate only tail behaviour, such subsets do not include enough information to identify their location. More promising is the development of reversible jump MCMC algorithms (Green (1995)), which have already been successfully applied in change-point applications, although not in the context of extremes.

Change-point methods are investigated here for a single time-series. However, in portfolio management it is often of interest to assess the likelihood of multiple stocks being extreme

at the same time. As for univariate extremes, such extreme dependence may change over time and abruptly due to exogenous events. A methodology to identify extreme dependence change-points using a Bayesian approach is the subject of ongoing research.

## Appendix A: MCMC algorithms

### A.1 CMGPD

Sampling is carried out in blocks with Metropolis–Hastings proposals. A parameter with a superscript  $(s)$  denotes its value at the  $s$ -th iteration of the algorithm. Let  $\mu = \{\mu_1, \dots, \mu_l\}$ ,  $\eta = \{\eta_1, \dots, \eta_l\}$ ,  $p = \{p_1, \dots, p_l\}$ ,  $u = \{u_1, \dots, u_k\}$ ,  $\sigma = \{\sigma_1, \dots, \sigma_k\}$ ,  $\xi = \{\xi_1, \dots, \xi_k\}$  and  $\tau = \{\tau_0, \dots, \tau_k\}$ . We denote  $\xi_{<j} = \{\xi_1, \dots, \xi_{j-1}\}$ ,  $\xi_{\geq j} = \{\xi_j, \dots, \xi_k\}$  and similarly for other parameters. Recall that  $\Phi = \{\mu, \eta, p\}$  and  $\Psi = \{\xi, \sigma, u\}$ . At each iteration  $s$ , parameters are updated as follows:

*Sampling  $\xi$ .* The proposal transition kernel for each  $\xi_j$ ,  $j \in [k]$ , where  $k$  is the total number of regimes, is given by a truncated Normal

$$N(\xi_j^{(s)}, V_{\xi_j}) \mathbb{1}_{(-\sigma_j^{(s)}(M_j^{(s)} - u_j^{(s)}), \infty)}$$

where  $V_{\xi_j}$  is a variance appropriately chosen to ensure chain mixing and  $M_j^{(s)}$  is the maximum of the observations in  $(\tau_{j-1}^{(s)}, \tau_j^{(s)})$ . So,  $\xi_j^{(s+1)} = \xi_j^*$  with probability  $\alpha_{\xi_j}$ , where

$$\alpha_{\xi_j} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_N(\xi_j^{(s)}, V_{\xi_j}) \mathbb{1}_{(-\sigma_j^{(s)}(M_j^{(s)} - u_j^{(s)}), \infty)}}{\pi(\tilde{\Theta}|x) f_N(\xi_j^*, V_{\xi_j}) \mathbb{1}_{(-\sigma_j^{(s)}(M_j^{(s)} - u_j^{(s)}), \infty)}} \right\},$$

$\Theta^* = \{\Phi^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^*, \tau^{(s)}\}$ ,  $\tilde{\Theta} = \{\Phi^{(s)}, u^{(s)}, \sigma^{(s)}, \xi_{<j}^{(s+1)}, \xi_{\geq j}^{(s)}, \tau^{(s)}\}$  and  $\xi^* = \{\xi_{<j}^{(s+1)}, \xi_j^*, \xi_{>j}^{(s)}\}$ .

*Sampling  $\sigma$ .* The proposal transition kernel for each  $\sigma_j$ ,  $j = 1 \in [k]$ , depends on the value of  $\xi_j^{(s+1)}$ . If  $\xi_j^{(s+1)} \geq 0$ , then  $\sigma_j^*$  is sampled from the Gamma distribution  $G(\sigma_j^{(s)}, \sigma_j^{(s)2} / V_{\sigma_j})$  where  $V_{\sigma_j}$  is the variance of the proposal distribution appropriately chosen to ensure chain mixing. If  $\xi_j^{(s+1)} < 0$ , then  $\sigma_j^*$  is sampled from a  $N(\sigma_j^{(s)}, V_{\sigma_j}) \mathbb{1}_{(-\xi_j^{(s+1)}(M_j^{(s)} - u_j^{(s)}), \infty)}$ . So,  $\sigma_j^{s+1} = \sigma_j^*$  with probability  $\alpha_{\sigma_j}$  where, if  $\xi_j^{(s+1)} < 0$ ,

$$\alpha_{\sigma_j} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_N(\sigma_j^{(s)}, V_{\sigma_j}) \mathbb{1}_{(-\xi_j^{(s+1)}(M_j^{(s)} - u_j^{(s)}), \infty)}}{\pi(\tilde{\Theta}|x) f_N(\sigma_j^*, V_{\sigma_j}) \mathbb{1}_{(-\xi_j^{(s+1)}(M_j^{(s)} - u_j^{(s)}), \infty)}} \right\},$$

and if  $\xi_j^{(s+1)} > 0$ ,

$$\alpha_{\sigma_j} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_G(\sigma_j^{(s)} | \sigma_j^*, \sigma_j^{*2} / V_{\sigma_j})}{\pi(\tilde{\Theta}|x) f_G(\sigma_j^{(s)} | \sigma_j^{(s)}, \sigma_j^{(s)2} / V_{\sigma_j})} \right\},$$

$\Theta^* = \{\Phi^{(s)}, u^{(s)}, \sigma^*, \xi^{(s+1)}, \tau^{(s)}\}$ ,  $\tilde{\Theta} = \{\Phi^{(s)}, u^{(s)}, \sigma_{<j}^{(s+1)}, \sigma_{\geq j}^{(s)}, \xi^{(s+1)}, \tau^{(s)}\}$  and  $\sigma^* = \{\sigma_{<j}^{(s+1)}, \sigma_j^*, \sigma_{>j}^{(s)}\}$ .

*Sampling  $u$ .* The thresholds  $u_j^*$  are sampled from a  $N(u_j^{(s)}, V_{u_j})\mathbb{1}_{(a_j^{(s+1)}, \infty)}$  distribution where  $a_j^{s+1}$  is the minimum of the observations in  $(\tau_{j-1}^{(s)}, \tau_j^{(s)})$  if  $\xi_j^{(s+1)} \geq 0$  and  $a_j^{s+1} = M_j^{(s)} + \sigma_j^{(s+1)}/\xi_j^{(s+1)}$  if  $\xi_j^{(s+1)} < 0$ . The lower limit of the truncation is chosen to satisfy the sample space of the GPD. The variance  $V_{u_j}$  is chosen to ensure appropriate chain mixing. So  $u_j^{(s+1)} = u_j^*$  with probability  $\alpha_{u_j}$ , where

$$\alpha_{u_j} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_N(u_j^{(s)}, V_{u_j}) \mathbb{1}_{(a_j^{(s+1)}, \infty)}}{\pi(\tilde{\Theta}|x) f_N(u_j^*, V_{u_j}) \mathbb{1}_{(a_j^{(s+1)}, \infty)}} \right\},$$

$\Theta^* = \{\Phi^{(s)}, u^*, \sigma^{(s+1)}, \xi^{(s+1)}, \tau^{(s)}\}$ ,  $\tilde{\Theta} = \{\Phi^{(s)}, u_{<j}^{(s+1)}, u_{\geq j}^{(s)}, \sigma^{(s+1)}, \xi^{(s+1)}, \tau^{(s)}\}$  and  $u^* = \{u_{<j}^{(s+1)}, u_j^*, u_{>j}^{(s)}\}$ .

*Sampling  $\eta$ .* The proposal kernel for  $\eta_z$ ,  $z \in [l]$ , where  $l$  is the number of mixture components, is taken as the Gamma distribution  $G(\eta_z^{(s)}, \eta_z^{(s)2}/V_{\eta_z})$ , where  $V_{\eta_z}$  is chosen to ensure appropriate chain mixing. So  $\eta_z^{(s+1)} = \eta_z^*$  with probability  $\alpha_{\eta_z}$ , where

$$\alpha_{\eta_z} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_G(\eta_z^{(s)}|\eta_z^*, \eta_z^{*2}/V_{\eta_z})}{\pi(\tilde{\Theta}|x) f_G(\eta_z^*|\eta_z^{(s)}, \eta_z^{(s)2}/V_{\eta_z})} \right\},$$

$\Theta^* = \{\mu^{(s)}, \eta^*, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}$ ,  $\tilde{\Theta} = \{\mu^{(s)}, \eta_{<z}^{(s+1)}, \eta_{\geq z}^{(s)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}$  and  $\eta^* = \{\eta_{<z}^{(s+1)}, \eta_z^*, \eta_{>z}^{(s)}\}$ .

*Sampling  $\mu$ .* The proposal kernel for  $\mu_z$ ,  $z \in [l]$ , is taken as the Gamma distribution  $G(\mu_z^{(s)}, \mu_z^{(s)2}/V_{\mu_z})\mathbb{1}_{(\mu_1^{(s+1)} < \dots < \mu_{z-1}^{(s+1)} < \mu_z^{(s)} < \dots < \mu_h^{(s)})}$  where  $V_{\mu_z}$  is chosen to ensure appropriate chain mixing. So  $\mu_z^{(s+1)} = \mu_z^*$  with probability  $\alpha_{\mu_z}$ , where

$$\alpha_{\mu_z} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_G(\mu_z^{(s)}|\mu_z^*, \mu_z^{*2}/V_{\mu_z}) \mathbb{1}_{(\mu_1^{(s+1)} < \dots < \mu_z^* < \dots < \mu_h^{(s)})}}{\pi(\tilde{\Theta}|x) f_G(\mu_z^*|\mu_z^{(s)}, \mu_z^{(s)2}/V_{\mu_z}) \mathbb{1}_{(\mu_1^{(s+1)} < \dots < \mu_z^{(s)} < \dots < \mu_h^{(s)})}} \right\},$$

$\Theta^* = \{\mu^*, \eta^{(s+1)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}$ ,  $\tilde{\Theta} = \{\mu_{<z}^{(s+1)}, \mu_{\geq z}^{(s)}, \eta^{(s+1)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}$  and  $\mu^* = \{\mu_{<z}^{(s+1)}, \mu_z^*, \mu_{>z}^{(s)}\}$ .

*Sampling  $p$ .* The vector of weights is proposed from a Dirichlet

$$D_h(V_p p_1^{(s)}, \dots, V_p p_h^{(s)}),$$

where  $V_p$  is chosen to ensure chain mixing. So,  $p^{(s+1)} = p^*$  with probability  $\alpha_p$ , where:

$$\alpha_p = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_D(p^{(s)}|p^*)}{\pi(\tilde{\Theta}|x) f_D(p^*|p^{(s)})} \right\},$$

$\Theta^* = \{\mu^{(s+1)}, \eta^{(s+1)}, p^*, \Psi^{(s+1)}, \tau^{(s)}\}$  and  $\tilde{\Theta} = \{\mu^{(s+1)}, \eta^{(s+1)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}$ .

*Sampling  $\tau$ .* The proposal transition kernel for each  $\tau_j$ ,  $j \in [k-1]$ , is given by a truncated Normal  $N(\tau_j^{(s)}, V_{\tau_j})\mathbb{1}_{(\tau_{j-1}^{(s+1)}, \tau_{j+1}^{(s)})}$ , where  $V_{\tau_j}$  is chosen to ensure chain mixing and the proposed value is rounded to the closest integer. The truncated Normal proposal allows for a straightforward implementation of our adaptive MCMC step. So,  $\tau_j^{(s+1)} = \tau_j^*$  with probability  $\alpha_{\tau_j}$ , where

$$\alpha_{\tau_j} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_N(\tau_j^{(s)}, V_{\tau_j}) \mathbb{1}_{(\tau_{j-1}^{(s+1)}, \tau_{j+1}^{(s)})}}{\pi(\tilde{\Theta}|x) f_N(\tau_j^*, V_{\tau_j}) \mathbb{1}_{(\tau_{j-1}^{(s+1)}, \tau_{j+1}^{(s)})}} \right\},$$

$$\Theta^* = \{\Phi^{(s+1)}, \Psi^{(s+1)}, \tau^*\}, \tilde{\Theta} = \{\Phi^{(s+1)}, \Psi^{(s+1)}, \tau_{<j}^{(s+1)}, \tau_{\geq j}^{(s)}\} \text{ and } \tau^* = \{\tau_{<j}^{(s+1)}, \tau_j^*, \tau_{>j}^{(s)}\}.$$

## A.2 CMNPD

The steps for the CMNPD are the same as for the CMGPD with the only difference that the parameters of the mixture of Normals now need to be estimated, i.e. the means  $\mu = \{\mu_1, \dots, \mu_l\}$  and the variances  $\delta = \{\delta_1, \dots, \delta_l\}$ . In this case  $\Phi = \{\mu, \delta, p\}$ . At each iteration  $s$ , the Normal parameters are updated as follows:

*Sampling  $\mu$ .* The proposal kernel for  $\mu_z$ ,  $z \in [l]$ , is taken as the Gamma distribution  $G(\mu_z^{(s)}, \mu_z^{(s)2} / V_{\mu_z})\mathbb{1}_{(\mu_1^{(s+1)} < \dots < \mu_{z-1}^{(s+1)} < \mu_z^{(s)} < \dots < \mu_h^{(s)})}$  where  $V_{\mu_z}$  is chosen to ensure appropriate chain mixing. So  $\mu_z^{(s+1)} = \mu_z^*$  with probability  $\alpha_{\mu_z}$ , where

$$\alpha_{\mu_z} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_G(\mu_z^{(s)} | \mu_z^*, \mu_z^{(s)2} / V_{\mu_z}) \mathbb{1}_{(\mu_1^{(s+1)} < \dots < \mu_z^* < \dots < \mu_h^{(s)})}}{\pi(\tilde{\Theta}|x) f_G(\mu_z^* | \mu_z^{(s)}, \mu_z^{(s)2} / V_{\mu_z}) \mathbb{1}_{(\mu_1^{(s+1)} < \dots < \mu_z^{(s)} < \dots < \mu_h^{(s)})}} \right\},$$

$$\Theta^* = \{\mu^*, \delta^{(s)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}, \tilde{\Theta} = \{\mu_{<z}^{(s+1)}, \mu_{\geq z}^{(s)}, \delta^{(s)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\} \text{ and } \mu^* = \{\mu_{<z}^{(s+1)}, \mu_z^*, \mu_{>z}^{(s)}\}.$$

*Sampling  $\delta$ .* The proposal kernel for  $\delta_z$ ,  $z \in [l]$ , is taken as the Gamma distribution  $G(\delta_z^{(s)}, \delta_z^{(s)2} / V_{\delta_z})$  where  $V_{\delta_z}$  is chosen to ensure appropriate chain mixing. So  $\delta_z^{(s+1)} = \delta_z^*$  with probability  $\alpha_{\delta_z}$ , where

$$\alpha_{\delta_z} = \min \left\{ 1, \frac{\pi(\Theta^*|x) f_G(\delta_z^{(s)} | \delta_z^*, \delta_z^{(s)2} / V_{\delta_z})}{\pi(\tilde{\Theta}|x) f_G(\delta_z^* | \delta_z^{(s)}, \delta_z^{(s)2} / V_{\delta_z})} \right\},$$

$$\Theta^* = \{\mu^{(s+1)}, \delta^*, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\}, \tilde{\Theta} = \{\mu^{(s+1)}, \delta_{<z}^{(s+1)}, \delta_{\geq z}^{(s)}, p^{(s)}, \Psi^{(s+1)}, \tau^{(s)}\} \text{ and } \delta^* = \{\delta_{<z}^{(s+1)}, \delta_z^*, \delta_{>z}^{(s)}\}.$$

## Appendix B: Additional simulations

In this section we report the model selection criteria from models estimated over additional simulated dataset as well as tail and changepoint parameter estimates from the true generating model. In all cases for the bulk we use the same setup as in Section 3:  $(\mu_1, \mu_2) = (2, 8)$   $(\eta_1, \eta_2) = (4, 8)$  and  $(p_1, p_2) = (2/3, 1/3)$ .

### B.1 All parameters varying, small dataset

Here 2500 observations were simulated using the following setup:  $(\xi_1, \xi_2, \xi_3) = (-0.4, 0, 0.4)$ ,  $(\sigma_1, \sigma_2, \sigma_3) = (0.5, 1, 1.5)$ ,  $(u_1, u_2, u_3) = (80th, 85th, 90th)$  percentiles and  $\tau = \{700, 1800\}$ .

Model selection criteria:

Model	BIC	DIC	WAIC
MGPD <sub>2</sub>	10,607.89	10,550.86	10,552.81
MGPD <sub>3</sub>	10,332.06	10,282.86	10,259.09
MGPD <sub>4</sub>	10,347.70	10,279.22	10,253.94
CMGPD <sub>2</sub> <sup>2</sup>	10,474.25	10,436.11	10,438.43
CMGPD <sub>2</sub> <sup>3</sup>	<b>10,297.16</b>	<b>10,162.71</b>	10,210.84
CMGPD <sub>2</sub> <sup>4</sup>	10,256.82	10,199.15	<b>10,212.05</b>

Parameter estimates:

$\tau_1 = 700$ 696 (685, 702)	$\tau_2 = 1800$ 1781 (1706, 1837)	
$\xi_1 = -0.4$ -0.47 (-0.50, -0.40)	$\xi_2 = 0$ -0.08 (-0.26, 0.18)	$\xi_3 = 0.4$ 0.36 (0.04, 1.66)
$\sigma_1 = 0.5$ 0.59 (0.53, 0.65)	$\sigma_2 = 1.0$ 1.10 (0.78, 1.37)	$\sigma_3 = 1.5$ 2.14 (0.72, 3.12)
$u_1 = 6.99$ 7.00 (6.98, 7.01)	$u_2 = 7.99$ 7.95 (7.84, 8.07)	$u_3 = 9.22$ 9.07 (7.01, 15.23)

### B.2 Only shape $\xi$ varying, large dataset

Here 5000 observations were simulated using the following setup:  $(\xi_1, \xi_2, \xi_3) = (-0.4, 0, 0.4)$ ,  $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, 1)$ ,  $(u_1, u_2, u_3) = (80th, 80th, 80th)$  percentiles and  $\tau = \{2000, 3500\}$ .

Model selection criteria:

Model	BIC	DIC	WAIC
MGPD <sub>2</sub>	20,301.36	20,240.71	20,241.5
MGPD <sub>3</sub>	20,321.17	20,247.41	20,240.51
MGPD <sub>4</sub>	20,348.74	20,240.86	20,239.99
CMGPD <sub>2</sub> <sup>2</sup>	20,152.07	20,097.79	20,099.32
CMGPD <sub>2</sub> <sup>3</sup>	<b>20,135.62</b>	<b>20,081.12</b>	<b>20,083.55</b>
CMGPD <sub>2</sub> <sup>4</sup>	20,143.16	20,083.68	20,085.73

Parameter estimates:

$\tau_1 = 2000$ 2185 (2083, 2244)	$\tau_2 = 3500$ 3524 (3203, 3662)	
$\xi_1 = -0.4$ -0.37 (-0.44, -0.28)	$\xi_2 = 0$ 0.14 (-0.01, 0.30)	$\xi_3 = 0.4$ 0.51 (0.36, 0.72)
$\sigma_1 = 1.0$ 0.95 (0.84, 1.05)	$\sigma_2 = 1.0$ 0.93 (0.75, 1.16)	$\sigma_3 = 1.0$ 0.90 (0.69, 1.08)
$u_1 = 6.99$ 7.01 (6.99, 7.08)	$u_2 = 6.99$ 7.00 (6.99, 7.01)	$u_3 = 6.99$ 7.00 (6.96, 7.03)

### B.3 Only shape $\xi$ varying, small dataset

Here 5000 observations were simulated using the following setup:  $(\xi_1, \xi_2, \xi_3) = (-0.4, 0, 0.4)$ ,  $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, 1)$ ,  $(u_1, u_2, u_3) = (80th, 80th, 80th)$  percentiles and  $\tau = \{700, 1800\}$ .

Model selection criteria:

Model	BIC	DIC	WAIC
MGPD <sub>2</sub>	10,188.85	10,129.77	10,131.18
MGPD <sub>3</sub>	10,191.81	10,103.08	10,118.27
MGPD <sub>4</sub>	10,209.31	10,104.83	10,116.84
CMGPD <sub>2</sub> <sup>2</sup>	10,148.59	10,076.94	10,098.06
CMGPD <sub>2</sub> <sup>3</sup>	<b>10,115.37</b>	<b>10,056.42</b>	<b>10,067.77</b>
CMGPD <sub>2</sub> <sup>4</sup>	10,118.97	10,063.40	10,070.98

Parameter estimates:

$\tau_1 = 700$ 712 (606, 753)	$\tau_2 = 1800$ 1747 (1512, 2027)	
$\xi_1 = -0.4$ -0.40 (-0.46, -0.29)	$\xi_2 = 0$ -0.08 (-0.22, 0.07)	$\xi_3 = 0.4$ 0.38 (0.17, 0.62)
$\sigma_1 = 1.0$ 1.04 (0.89, 1.20)	$\sigma_2 = 1.0$ 1.06 (0.88, 1.30)	$\sigma_3 = 1.0$ 1.08 (0.81, 1.49)
$u_1 = 6.99$ 7.00 (6.97, 7.03)	$u_2 = 6.99$ 6.99 (6.93, 7.08)	$u_3 = 6.99$ 7.00 (6.99, 7.03)

### B.4 Only scale $\sigma$ varying, large dataset

Here 5000 observations were simulated using the following setup:  $(\xi_1, \xi_2, \xi_3) = (-0.4, -0.4, -0.4)$ ,  $(\sigma_1, \sigma_2, \sigma_3) = (0.5, 1, 1.5)$ ,  $(u_1, u_2, u_3) = (80th, 80th, 80th)$  percentiles and  $\tau = \{2000, 3500\}$ .

Model selection criteria:

Model	BIC	DIC	WAIC
MGPD <sub>2</sub>	19,546.09	19,474.53	19,479.49
MGPD <sub>3</sub>	19,577.26	19,475.03	19,482.27
MGPD <sub>4</sub>	19,598.41	19,475.55	19,480.70
CMGPD <sub>2</sub> <sup>2</sup>	19,214.03	19,161.32	19,162.84
CMGPD <sub>2</sub> <sup>3</sup>	<b>19,162.44</b>	<b>19,100.29</b>	<b>19,106.18</b>
CMGPD <sub>2</sub> <sup>4</sup>	19,175.15	19,104.49	19,111.33

Parameter estimates:

$\tau_1 = 2000$ 2000 (1967, 2019)	$\tau_2 = 3500$ 3490 (3444, 3509)	
$\xi_1 = -0.4$ -0.37 (-0.43, -0.28)	$\xi_2 = -0.4$ -0.47 (-0.50, -0.38)	$\xi_3 = -0.4$ -0.40 (-0.47, -0.29)
$\sigma_1 = 0.5$ 0.47 (0.42, 0.51)	$\sigma_2 = 1.0$ 1.10 (0.97, 1.19)	$\sigma_3 = 1.5$ 1.47 (1.25, 1.67)
$u_1 = 6.99$ 7.00 (6.99, 7.00)	$u_2 = 6.99$ 7.00 (6.99, 7.00)	$u_3 = 6.99$ 7.05 (6.99, 7.20)

### B.5 Only scale $\sigma$ varying, small dataset

Here 5000 observations were simulated using the following setup:  $(\xi_1, \xi_2, \xi_3) = (-0.4, -0.4, -0.4)$ ,  $(\sigma_1, \sigma_2, \sigma_3) = (0.5, 1, 1.5)$ ,  $(u_1, u_2, u_3) = (80th, 80th, 80th)$  percentiles and  $\tau = \{700, 1800\}$ .

Model selection criteria:

Model	BIC	DIC	WAIC
MGPD <sub>2</sub>	9897.29	9838.18	9840.41
MGPD <sub>3</sub>	9815.92	9732.19	9736.84
MGPD <sub>4</sub>	9838.44	9730.86	9737.25
CMGPD <sub>2</sub> <sup>2</sup>	9811.32	9749.76	9766.44
CMGPD <sub>2</sub> <sup>3</sup>	<b>9740.93</b>	<b>9684.35</b>	<b>9698.31</b>
CMGPD <sub>2</sub> <sup>4</sup>	9755.60	9685.52	9700.05

Parameter estimates:

$\tau_1 = 700$ 729 (663, 752)	$\tau_2 = 1800$ 1848 (1747, 1885)	
$\xi_1 = -0.4$ -0.47 (-0.50, -0.40)	$\xi_2 = -0.4$ -0.46 (-0.50, -0.38)	$\xi_3 = -0.4$ -0.40 (-0.50, -0.25)
$\sigma_1 = 0.5$ 0.58 (0.52, 0.64)	$\sigma_2 = 1.0$ 1.05 (0.92, 1.17)	$\sigma_3 = 1.5$ 1.60 (1.25, 1.92)
$u_1 = 6.99$ 6.96 (6.93, 7.00)	$u_2 = 6.99$ 6.98 (6.94, 7.01)	$u_3 = 6.99$ 6.91 (6.67, 7.19)

## References

- Albert, J. H. and Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* **11**, 1–15.
- Behrens, C. N., Lopes, H. F. and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling* **4**, 227–244. MR2062102 <https://doi.org/10.1191/1471082X04st0750a>
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**, 542–547.
- Caldara, D., Fuentes-Albero, C., Gilchrist, S. and Zakrajšek, E. (2016). The macroeconomic impact of financial and uncertainty shocks. *European Economic Review* **88**, 185–207.
- Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society Series C* **41**, 389–405. MR1278223
- Carreau, J. and Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: The univariate case. *Extremes* **12**, 53–76. MR2480723 <https://doi.org/10.1007/s10687-008-0068-0>
- Castellanos, M. E. and Cabras, S. (2007). A default Bayesian procedure for the generalized Pareto distribution. *Journal of Statistical Planning and Inference* **137**, 473–483. MR2298951 <https://doi.org/10.1016/j.jspi.2006.01.006>
- Catania, L. and Bernardi, M. (2017). MCS: Model confidence set procedure. R package version 0.1.3. Available at <https://CRAN.R-project.org/package=MCS>.
- Chang, C., Jiménez-Martín, J., McAleer, M. and Pérez-Amaral, T. (2011). Risk management of risk under the Basel accord: Forecasting value-at-risk of VIX futures. *Managerial Finance* **37**, 1088–1106.
- Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics* **86**, 221–241. MR1649222 [https://doi.org/10.1016/S0304-4076\(97\)00115-2](https://doi.org/10.1016/S0304-4076(97)00115-2)
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B* **52**, 393–442. MR1086795
- de Carvalho, M., Pereira, S., Pereira, P. and de Zea Bermudez, Z. (2020). An extreme value Bayesian lasso for the conditional bulk and tail. Available at [arXiv:201007164](https://arxiv.org/abs/201007164).
- Dierckx, G. and Teugels, J. L. (2010). Change point analysis of extreme values. *EnvironMetrics* **21**, 661–686. MR2838438 <https://doi.org/10.1002/env.1041>
- DuMouchel, W. H. (1983). Estimating the stable index  $\alpha$  in order to measure tail thickness: A critique. *The Annals of Statistics* **11**, 1019–1031. MR0865342 <https://doi.org/10.1214/aos/1176346318>
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall. MR2260716
- Gilli, M. and Kellezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics* **27**, 207–228.
- González-Rivera, G., Lee, T. and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* **20**, 629–645.
- Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**, 27–62.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732. MR1380810 <https://doi.org/10.1093/biomet/82.4.711>
- Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set. *Econometrica* **79**, 453–497. MR2809377 <https://doi.org/10.3982/ECTA5771>
- Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *Test* **23**, 219–255. MR3210268 <https://doi.org/10.1007/s11749-014-0368-4>
- Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics* **14**, 285–299. MR2405331 <https://doi.org/10.1007/s10651-007-0014-3>
- Jarušková, D. and Rencová, M. (2008). Analysis of annual maximal and minimal temperatures for some European cities by change point methods. *EnvironMetrics* **19**, 221–233. MR2420467 <https://doi.org/10.1002/env.865>
- Killick, R. and Eckley, I. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software* **58**, 1–19.
- Ko, S. I. M., Chong, T. T. L. and Ghosh, P. (2015). Dirichlet process hidden Markov multiple change-point model. *Bayesian Analysis* **10**, 275–296. MR3420883 <https://doi.org/10.1214/14-BA910>
- Leonelli, M. and Gamerman, D. (2020). Semiparametric bivariate modelling with flexible extremal dependence. *Statistics and Computing* **30**, 221–236. MR4064619 <https://doi.org/10.1007/s11222-019-09878-w>
- Lima, S. R., Nascimento, F. F. and Ferraz, V. R. S. (2018). Regression models for time-varying extremes. *Journal of Statistical Computation and Simulation* **88**, 235–249. MR3740727 <https://doi.org/10.1080/00949655.2017.1385788>

- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M. and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis* **55**, 2137–2157. MR2785120 <https://doi.org/10.1016/j.csda.2011.01.005>
- McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* **7**, 271–300.
- Nascimento, F. F., Gamerman, D. and Davis, R. (2016). A Bayesian semi-parametric approach to extreme regime identification. *Brazilian Journal of Probability and Statistics* **30**, 540–561. MR3582389 <https://doi.org/10.1214/15-BJPS293>
- Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2011). Regression models for exceedance data via the full likelihood. *Environmental and Ecological Statistics* **18**, 495–512. MR2832907 <https://doi.org/10.1007/s10651-010-0148-6>
- Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2012). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing* **22**, 661–675. MR2865043 <https://doi.org/10.1007/s11222-011-9270-z>
- Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2016). Time-varying extreme pattern with dynamic models. *Test* **25**, 131–149. MR3463806 <https://doi.org/10.1007/s11749-015-0444-4>
- Nascimento, F. F. and Moura e Silva, W. V. (2017). A Bayesian model for multiple change point to extremes, with application to environmental and financial data. *Journal of Applied Statistics* **44**, 2410–2426. MR3683052 <https://doi.org/10.1080/02664763.2016.1254733>
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131. MR0423667
- Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**, 349–367. MR2749836 <https://doi.org/10.1198/jcgs.2009.06134>
- Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys* **28**, 82–108.
- Sampid, M. G., Hasim, H. M. and Dai, H. (2018). Refining value-at-risk estimates using a Bayesian Markov-switching GJR-GARCH copula-EVT model. *PLoS ONE* **13**, e0198753.
- Samuel, Y. M. Z. (2008). Value at risk and conditional extreme value theory via Markov regime switching models. *Journal of Futures Markets: Futures, Options, and Other Derivative Products* **28**, 155–181.
- Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Statistical Journal* **10**, 33–60. MR2912370
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639. MR1979380 <https://doi.org/10.1111/1467-9868.00353>
- Stephens, D. A. (1994). Bayesian retrospective multiple-changepoint identification. *Journal of the Royal Statistical Society Series C* **43**, 159–178.
- Tancredi, A., Anderson, C. and O’Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes* **9**, 87–106. MR2329799 <https://doi.org/10.1007/s10687-006-0009-8>
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594. MR2756194
- Wiper, M., Riouins Insua, D. and Ruggeri, F. (2001). Mixtures of Gamma distributions with applications. *Journal of Computational and Graphical Statistics* **10**, 440–454. MR1939034 <https://doi.org/10.1198/106186001317115054>
- Zhao, X., Scarrott, C. J., Oxley, L. and Reale, M. (2011). GARCH dependence in extreme value models with Bayesian inference. *Mathematics and Computers in Simulation* **81**, 1430–1440. MR2781673 <https://doi.org/10.1016/j.matcom.2010.08.002>